PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G10L 5/00

A2

(11) International Publication Number: WO 99/40569

(43) International Publication Date: 12 August 1999 (12.08.99)

(21) International Application Number: PCT/FI99/00097

(22) International Filing Date: 9 February 1999 (09.02.99)

(30) Priority Data:
980298 9 February 1998 (09.02.98) FI

(71) Applicant *(for all designated States except US)*: NOKIA TELECOMMUNICATIONS OY [FI/FI]; P.O. Box 300, FIN–00045 Nokia Group (FI).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: KAPANEN, Pekka [FI/FI]; Näyttelijänkatu 21 E 14, FIN–33720 Tampere (FI). VAINIO, Janne [FI/FI]; Laurintie 16 C, FIN–33880 Lempäälä (FI).

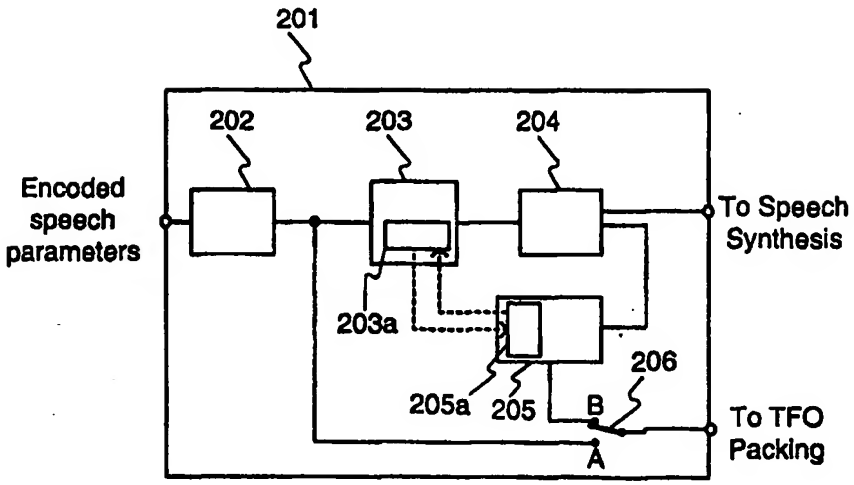
(74) Agent: BERGGREN OY AB; P.O. Box 16, FIN–00101 Helsinki (FI).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: A DECODING METHOD, SPEECH CODING PROCESSING UNIT AND A NETWORK ELEMENT



(57) Abstract

This invention is related to tandem free operation (TFO) in mobile cellular systems. The present invention implements a tandem free operation by using a special feedback loop which makes the decoded parameters available, performs the comfort noise insertion and bad frame handling operations, produces the parameter quantisation indices corresponding to the output of these operations, and synchronises the speech encoders and the speech decoders in the transmission path from the uplink mobile station to the downlink mobile station. This functionality is realized by partly decoding and re–encoding the parameters and synchronising and resetting the quantiser prediction memories in a specific manner. A basic idea of the invention is, that during BFH and CNI processes, a re–encoding block produces models of encoded speech parameters from the BFH/CNI processed speech parameters. These models of encoded speech parameters are then transmitted to the receiving end. The present invention provides a solution to the problem created by predictive, more generally non–stateless encoders in TFO operation.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

A decoding method, speech coding processing unit and a network element

TECHNICAL FIELD OF THE INVENTION

This invention is related to tandem free operation (TFO) in mobile cellular systems. The invention is further related to that, which is stated in the preamble of Claim 1.

BACKGROUND OF THE INVENTION

For convenience, various abbreviations used in this specification are presented here:

TFO	Tandem Free Operation
CNI	Comfort Noise Insertion
CN	Comfort Noise
BFH	Bad Frame Handling
UMS	Uplink Mobile Station
DMS	Downlink Mobile Station
UBS	Uplink Base Station
UTR	Uplink Transcoder
DTR	Downlink Transcoder
DBS	Downlink Base Station
AI	Air Interface
PCM	Pulse Coded Modulation
PSTN	Public Switched Telephone Network
UAI	Uplink Air Interface
DAI	Downlink Air Interface
DTX	Discontinuous Transmission
VAD	Voice Activity Detection

Speech frames received by the mobile network from a mobile communication means can be roughly classified into three classes: a) uncorrupted, i.e. good speech frames; b) corrupted speech frames; and c) frames generated during discontinued transmission (DTX) mode, which frames generally include silence descriptor (SID) frames and unusable frames received during the transmission pause.

In normal mode of operation, a mobile unit encodes the speech to be transmitted, and the encoded speech is decoded after transmission through the air interface. When a mobile unit receives a call, the speech is encoded at the network side of the

air interface, and decoded in the receiving mobile unit. Therefore, in normal mode of operation without special arrangements taking place, speech is encoded and decoded twice in a mobile-to-mobile call, resulting in a decrease of perceived speech quality. Tandem free operation (TFO) is a mode of operation between two mobile units, in which the speech is encoded only once, and the speech is transmitted in the encoded form over the network to the receiving mobile unit.

Since it is not feasible to send the error indication information contained in erroneous frames and the side information contained in DTX frames through the mobile network to the receiving end, it has been found feasible in GSM to transmit during TFO operation all frames over the A-interface as good frames. The A-interface is the interface between the transmitting and receiving mobile networks. In conventional non-TFO operation, the speech is transmitted over the A-interface as a digital real-time waveform as PCM-coded samples.

A so-called bad frame handling procedure is used in converting erroneous frames received from the mobile communication means to good frames for transmission over the A-interface. In order to send comfort noise information contained in DTX frames over the A-interface, the comfort noise information has to be converted into good speech frames for transmission over the A-interface.

Comfort noise insertion is discussed first in more detail in the following paragraphs, then bad frame handling.

Comfort Noise Insertion

In Discontinuous Transmission (DTX), a Voice Activity Detector (VAD) detects on the transmit side whether or not the user is speaking. When the user is speaking, speech parameters descriptive of the input speech are produced in the speech encoder for each frame and transmitted to the receiving end. However, when the user stops speaking, parameters descriptive of the prevailing background noise are produced and transmitted to the receive side instead of the speech parameters. After this, the transmission is switched off. The transmission is resumed at the normal transmission rate when the user starts speaking again, or at a low rate to update the parameters describing the background noise while the user does not speak in order to adapt to changes occurring in the prevailing background noise during the transmission pause. Throughout this text, these parameters describing the prevailing background noise are referred to as comfort noise parameters or CN parameters.

At the receiving end, speech is synthesised whenever good speech parameter frames are received. However, when comfort noise parameters have been received, after which the transmission has been switched off, the speech decoder uses the received comfort noise parameters to locally synthesise noise with characteristics similar to the background noise on the transmit side. This synthetic noise is commonly referred to as Comfort Noise (CN), and the procedure of generating CN locally on the receive side is commonly referred to as Comfort Noise Insertion (CNI).

The updated comfort noise parameters are applied to the CNI procedure either immediately when received, or by gradually interpolating frame-by-frame from the previously received comfort noise parameter values to the updated parameter values. The former method guarantees that the comfort noise parameters are always as fresh as possible. However, the former method may result in stepwise effects in the perceived CN characteristics, and thus the latter method of interpolation is often used to alleviate this inconvenience. The latter method has the drawback in that the interpolation of the received comfort noise parameters introduces some delay in characterisation of the prevailing background noise, thereby introducing some contrast between the actual background noise and the CN.

Details of comfort noise insertion are described in the ETSI specification ETS 300 580-4, "European digital cellular telecommunications system (Phase 2); Comfort noise aspect for full rate speech traffic channels (GSM 06.12)", September 1994, which is hereinafter called the GSM 06.12 specification.

Bad Frame Handling

Bad frame handling (BFH) refers to a substitution procedure for frames containing errors. The purpose of the frame substitution is to conceal the effect of corrupted frames, since normal decoding of corrupted or lost speech frames would result in very unpleasant noise effects. In order to improve the subjective quality of the received speech, the first lost speech frame is substituted with either a repetition or an extrapolation of the previous good speech frames. Corrupted speech frames are not transmitted to the receiving end. If a number of consecutive frames is lost, the output of the speech decoder is gradually muted in order to indicate the user about the problems in the connection. The frame substitution procedure is discussed in the ETSI specification draft pr ETS 300 580-3, "Digital cellular telecommunications system; Full rate speech; Part 3: Substitution and muting of lost frames for full rate

speech channels (GSM 06.11 version 4.0.5)", November 1997, which is hereinafter called the GSM 06.11 specification.

Mobile To Mobile Calls

In the following, the flow of the speech data during a normal, non-TFO connection is discussed. The case of TFO operation is discussed after that.

The basic block diagram of the mobile to mobile call is illustrated in Figure 1. In an Uplink Mobile Station (UMS) 100, i.e. the mobile station in the transmitting end, the time-domain waveform is first divided into fixed-length frames and speech encoded in a speech coding block 101, i.e., transformed to speech coding parameters, which are then channel encoded in a channel coding block 102 by inserting redundant information for error correction purposes. These protected speech frames are then transmitted over the air interface (AI).

In an Uplink Base Station (UBS) 110, the channel decoding is performed in the channel decoding block 111, i.e., the channel errors are corrected and the redundant information is removed from the speech coding parameters. The speech coding parameters are transmitted through a serial Uplink Abis interface to an Uplink Transcoder (UTR) 120, where the speech coding parameters are transformed to a digital time-domain speech waveform in a speech decoding block 122. In normal non-TFO mode, the switch 121 is open as shown in figure 1, and the speech waveform is passed through a TFO packing block 123 essentially unchanged. The output of the UTR is transmitted through the A-interface to a public switched telephone network (PSTN) or to another mobile telephone network.

In a Downlink Transcoder (DTR) 130, the time-domain waveform is received from the A-interface. In non-TFO-operation, the switch 133 connects the output of the speech encoding block 132 to the output of the DTR, and the TFO extracting block 131 passes through the time-domain waveform unchanged. The waveform is transformed to speech coding parameters in the speech encoding block 132. The speech coding parameters are forwarded to the Downlink Abis interface.

In the downlink base station (DBS) 140, the speech parameters received from the Downlink Abis interface are channel encoded in the channel encoding block 141. The channel encoded parameters are transmitted to a Downlink Mobile Station (DMS) 150, i.e. the receiving mobile station. In the DMS, the channel coding is

removed in a channel decoding block 151 and the speech coding parameters are transformed back to a time-domain waveform, i.e. decoded speech, in the speech decoding block 152.

The problem in the conventional mode described above is the negative effect of two consecutive encodings on the quality of the transmitted speech signal. Since the encoding of the waveform in the speech encoding block 132 of the Downlink Transcoder (DTR) 130 is the second successive compression to the original input signal, the parameters in the output of the speech encoder 132 of the DTR 130 represent a time-domain waveform which is not a very accurate reproduction of the original speech waveform due to the errors created in two compressions. The tandem-free operation (TFO) was designed to alleviate this problem in at least some cases.

Tandem-Free Operation

In a mobile station to mobile station telephone call utilising a tandem-free mode of operation, hereinafter referred to as TFO, speech is transmitted by sending the parameters representing the time-domain speech waveform from an uplink mobile station speech encoder directly to a downlink mobile station speech decoder, without converting the parameters into a time-domain speech waveform in between the uplink transcoder and the downlink transcoder.

This significantly improves the speech quality because without TFO, the original speech signal is coded twice with the lossy speech compression algorithm which degrades the speech quality each time the compression is applied. The difference between the single encoding and the tandem encoding becomes even more important when the bit-rate of a speech codec is very low. The old high bit-rate speech coding standards, as exemplified by the G.711 standard of 64 kbit/s PCM coding, are very robust to successive coding. However, the state of the art speech coders operating in a range of 4 kbit/s to 16 kbit/s are quite sensitive to more than one successive coding.

The tandem-free operation according to prior art is discussed in the following with reference to Figure 1. In tandem-free operation, the speech parameters received by the speech decoding block 122 of the uplink transcoder 120 are embedded into the least significant bits of the decoded speech waveform in the TFO packing block 123, which is indicated in figure 1 by the closed position of the switch 121. The

speech waveform with the embedded speech parameters is then forwarded to the A-interface.

In order to enable the TFO mode, the downlink end of the call must naturally be in a mobile telephone network using the same speech coding standard as the uplink end. However, the call may be forwarded from the A-interface through several digital transmission links to the downlink mobile telephone network.

In the receiving end, the speech waveform with the embedded speech parameters is received from the A-interface by the downlink transcoder 130. The TFO extracting block 131 extracts the embedded speech parameters from the speech waveform. In TFO operation, the switch 133 connects the output of the TFO extracting block to the output of the downlink transcoder. The extracted original parameters are then forwarded to the downlink Abis interface and further via the downlink base station 140 through the air interface to the downlink mobile station, whose speech decoding block 152 then decodes the original speech parameters as encoded by the speech encoding block of the uplink mobile station 100.

Sometimes there are detected and undetected errors in the Air interface. These errors and the BFH operations can cause some mismatch between the parameters of speech encoder 101 of the transmitting mobile station and speech decoder 152 of the receiving mobile station. Usually these mismatches are diminished after the correct parameters have been received for several consecutive frames.

BFH and CNI handling in tandem free operation

Usually the functionality for bad frame handling and comfort noise insertion in the transmitting end is located in the speech decoder block 122 of the uplink transcoder 120. These functions are not illustrated in Figure 1. When any speech frames are corrupted or lost, or DTX transmission pauses occur, the speech decoder block 122 generates speech coding parameters corresponding to these situations as described previously.

As can be observed from Figure 1, the UMS 100, UBS 110, DBS 140 and the DMS 150 are not involved in the TFO operations concerning the BFH and CNI, but operate transparently as in the non-TFO case. The speech encoder 132 of the DTR operates normally during TFO as well, except that its output is not forwarded to the downlink Abis interface, but is replaced with the speech coding parameters

extracted from the A-interface stream instead. The operations concerning the BFH and CNI take place in the speech decoder 122 of the UTR 120.

A more detailed block diagram of the prior art speech decoder 122 realizing the CNI and BFH functions is shown in Figure 2. The encoded speech parameters, i.e. the parameter quantisation indices are extracted from the received information stream in parameter extracting blocks 122a. The BFH and CNI operations are performed on these parameter quantisation indices in BFI/CNI blocks 122b prior to the dequantisation (decoding) of the indices in dequantisation blocks 122c. After dequantisation, the parameters are used in speech synthesis in a speech synthesis block 122d to produce the decoded output signal. The BFI and CNI flags are signals produced by the uplink base station 110, which signals inform the decoder 122 about corrupted and DTX frames. The BFI/CNI blocks 122b are controlled by the BFI and CNI flags.

A similar block diagram with prior art TFO functionality is shown in Figure 3, which shows a diagram of the speech decoder 122 of an UTR 120 as well as the TFO packing block 123. As can be observed from Figure 3, the CNI and the BFH operations are performed on the parameter quantisation indices in the speech decoder 122. Therefore the tandem free operations in the UTR 120 are simply effected by packing (embedding) of the already available parameters from the decoder 122 into the time-domain waveform signal.

BFH operations during tandem free operation are straightforward, and can be effected in the same way as in non-TFO mode. The GSM 06.11 specification contains an example prior art solution of the BFH functionality, which can also be used during tandem free operation. The CNI operations are simple because the quantisations are memoryless, which means that all information during comfort noise generation or in the transitions between active speech and comfort noise is contained in the currently transmitted parameters. There are no problems for example in the resetting of the different parts of the transmission path. The prior art CNI solution is described in the specification GSM 06.12.

In tandem free operation, the parameter information packed to the signal transmitted to the A-interface must include all information needed to produce good speech frames, since the downlink mobile station is not aware of the CNI-operation at the uplink end. Due to this requirement, a simple conversion is performed on the comfort noise parameters to convert them to speech parameter frames. This involves

storing the most recent comfort noise parameters, and repeatedly forwarding them to the A-interface stream until updated comfort noise parameters are received and stored, or until active speech parameters are received. In case comfort noise parameter interpolation is desired as discussed earlier, this interpolation can be performed prior to forwarding the parameters to A-interface stream. Since comfort noise parameters do not include all parameters present in a good speech parameter frame, these missing speech parameters need to be created in some way during the conversion process.

Problems inherent in the prior art solutions

Figure 3 shows a decoder using conventional non-predictive quantisers. When the quantisers of the decoder are non-predictive as in Figure 3, BFH and CNI processing of the parameters do not create any problems. However, it is predictive quantisers that are used in the state of the art low rate encoders and decoders.

In a state of the art speech codec employing predictive quantisers, comfort noise insertion and bad frame handling operations have to be performed using the dequantised (decoded) parameters in the speech decoder, i.e. after the dequantiser blocks 122c and not before them as shown in Figure 3. The reason for this is that in predictive quantising and dequantising, the quantised entities (in this case, speech parameters) are not independent. When evaluating (decoding) predictively quantised entities, the evaluation result for each evaluated entity does not depend only on the quantised entity under evaluation, but also on the previous entities. Therefore, simple substitution of corrupted encoded parameters to suitable CN or BFH parameters is not possible. The substitution would have to adjust the substituting CN or BFH parameters according to the previously received good parameters, but since there is no knowledge of the development of the signal during the transmission pause or disturbance, the next good parameters received would depend on another history than that generated in the decoder, resulting in very annoying sound artifacts at the end of the pause. Therefore, CNI and BFH operations are effected after predictive dequantization on the decoded speech parameters, and coded speech parameters corresponding to CNI or BFH blocks are not available. Since the coded parameters describing CNI or BFH blocks are not available, they cannot be embedded in the time-domain speech waveform along with the rest of the coded parameters. Because of this problem, CNI and BFH operations are not possible in prior art tandem free operation, when the uplink mobile station uses a speech codec with predictive quantisers.

SUMMARY OF THE INVENTION

The object of the invention is to realize a method for implementing CNI and BFH operations in tandem free operation with predictively quantized speech parameters. A further object of the invention is to realize a speech decoder capable of CNI and BFH operations in connection with decoding of predictively quantized speech data in tandem free operation.

The objects are reached by producing re-encoded speech parameters from the dequantised BFH/CNI processed speech parameters, and transmitting these re-encoded parameters to the receiving end during BFH and CNI procedures.

The method according to the invention is characterized by that, which is specified in the characterizing part of the independent method claim. The speech coding processing unit according to the invention is characterized by that, which is specified in the characterizing part of the independent claim directed to a speech coding processing unit. The telecommunications network element according to the invention is characterized by that, which is specified in the characterizing part of the independent claim directed to a telecommunications network element. The dependent claims describe further advantageous embodiments of the invention.

The present invention implements a tandem free operation by using a special feedback loop which makes the decoded parameters available, performs the comfort noise insertion and bad frame handling operations, produces the parameter quantisation indices corresponding to the output of these operations, and synchronises the speech encoders and the speech decoders in the transmission path from the uplink mobile station to the downlink mobile station. This functionality is realized by partly decoding and re-encoding the parameters and synchronising and resetting the quantiser prediction memories in a specific manner. The present invention provides a solution to the problem created by predictive, more generally non-stateless encoders in TFO operation.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is described in more detail in the following with reference to the accompanying drawings, of which

Figure 1 illustrates the data flow of a mobile to mobile call according to prior art,

Figure 2 shows a block diagram of a prior art speech decoder,

Figure 3 shows a block diagram of a prior art speech decoder with TFO and CNI/BFH functionality,

Figure 4 shows a block diagram of a network element according to an advantageous embodiment of the invention,

Figure 5 shows a block diagram of a speech coding processing block of a speech speech coding processing unit according to an advantageous embodiment of the invention, and

Figure 6 shows a flow diagram of a method according to an advantageous embodiment of the invention.

Same reference numerals are used for similar entities in the figures.

DETAILED DESCRIPTION

A block diagram of a network element 220 such as, for example, an uplink transcoder or a speech coding processing unit, according to an advantageous embodiment of the invention is presented in Figure 4, and a speech coding processing block 201 according to an advantageous embodiment of the invention is presented in Figure 5. As can be observed from Figure 4, the network element comprises a speech decoder 200 and a TFO packing block 123. The network element receives from other elements located before the element in the transmission path encoded speech parameters and signals, such as a BFI flag and a CNI flag indicating various breaks in the signal flow, and produces an output signal comprising a time domain speech signal and, optionally, embedded encoded speech parameters. Further, in this embodiment the functions of blocks 122a, 122b, 122c of the prior art decoder are realized in speech coding processing blocks 201 according to the present invention. Such a speech coding processing block 201 is illustrated in Figure 5. In this exemplary embodiment, the outputs, inputs and the speech synthesis block 122d are similar to those of a prior art decoder 122 described previously, and are not described here in further detail. The speech coding processing block 201 comprises a parameter extraction block 202, a predictive

dequantiser block 203, a BFH/CNI processing block 204 and a predictive quantiser block 205. The dequantiser and quantiser blocks further have memories 203a, 205a.

The operation of a single speech coding processing block 201 according to Figure 5 is discussed in the following. Normal TFO operation is described first, i.e. operation between DTX pauses when no frames are corrupted in the uplink AI, secondly bad frame handling during TFO operation, and finally comfort noise insertion during TFO operation.

Normal TFO operation

In normal TFO operation, a parameter extraction block 202 extracts the desired parameters from the incoming frames of encoded speech parameters. The extracted encoded parameters are forwarded to a predictive dequantiser block 203, which dequantises the encoded parameters using information about previous dequantised parameters stored in the memory 203a of the dequantiser block 203. The dequantized parameters are forwarded to a BFH/CNI processing block 204, which in normal TFO operation forwards the parameters unchanged to speech synthesis. The extracted parameters from the parameter extraction block 202 are forwarded to TFO packing, which is represented by the position A of the switch member 206. In the present invention, an additional purpose of the decoding process is to provide correct initial values for the re-encoding quantiser block 205 memory for bad frame handling and discontinuous transmission operation.

BFH operation during TFO

In transitions from normal speech parameter transmission to BFH, the contents of the speech parameter dequantiser block memory 203a are copied to the quantiser memory 205a for proper initialisation of the re-encoding. This is represented by the arrow going from memory 203a to memory 205a.

In BFH operation, the BFH process is carried out on the decoded speech parameters produced by the predictive dequantiser block 203. The processed parameters are forwarded from the BFH/CNI processing block 204 to speech synthesis, and to the predictive quantiser block 205. The predictive quantiser block 205 re-encodes the dequantised and processed parameters to create new parameter quantisation indices and quantised parameters. The newly created re-quantised parameters are forwarded to TFO packing for transmission to the downlink end, which is represented by the

position B of the switch member 206. Thereafter the contents of the quantisation memory 205a are copied to the memory 203a of the dequantiser block 203. The copying operation is represented by the dashed arrow going from memory 205a to memory 203a in Figure 5. This copying operation results in the same state of the predictive dequantiser block 203, which would result, if the encoded parameters created by the quantising block 205 would in fact have been received from the uplink mobile station. Since the encoded parameters created by the quantising block 205 are forwarded via the TFO packing operation to the downlink mobile station, the speech decoder 200 of the UTR and the speech decoder 152 of the DMS are kept in synchronization.

CNI operation during TFO

In transitions from normal speech parameter transmission to DTX, the contents of the speech parameter dequantiser block memory 203a are copied to the quantiser memory 205a for proper initialisation of the re-encoding. This is represented by the arrow going from memory 203a to memory 205a.

In discontinuous transmission (DTX) mode of operation, the predictive quantisation can not be performed in the usual manner by updating the quantiser memories in each frame. Therefore, the synchronisation of the quantiser memories must be ensured between the encoder of UMS and the decoder of UTR with special arrangements to allow quantisation of the comfort noise parameters. The solution used in the prior art GSM system can be presented as an example of a suitable synchronisation method. According to GSM specification of enhanced full rate (EFR) coding during DTX mode, the quantiser memories are synchronised between the mobile unit and the transcoder by freezing the memories to identical values in both the encoder and the decoder for quantisation of the comfort noise parameters. This synchronization is described in further detail in the ETSI specification EN 301 247 V4.0.1 (November 1997) "Digital cellular telecommunications system (Phase 2); Comfort noise aspects for Enhanced Full Rate (EFR) speech traffic channels", also known as GSM specification 06.62 version 4.0.1. However, the present invention is not limited to the example of the GSM system. Any other mechanisms for synchronising the quantiser memories between the encoder of UMS and the decoder of UTR can be used as well in various embodiments of the invention.

In DTX operation the comfort noise parameters are transmitted from the UMS encoder to the UTR decoder and decoded using the special arrangements described in the previous paragraph. In each frame during DTX, the following steps are performed. The comfort noise parameters are either repeated or interpolated, as described previously in connection with prior art CNI operation. After the decoding operation, the parameters are re-encoded using the predictive quantiser block 205 as in the BFH case, and the memory 205a of the quantiser block 205 is updated. The newly created re-quantised parameters are forwarded to TFO packing for transmission to the downlink end. In this way the speech decoder 200 of the UTR and the speech decoder 152 of the DMS are kept in synchronization, since the encoded parameters created by the quantising block 205 are forwarded via the TFO packing operation to the downlink mobile station.

When the transmission of normal speech frames is resumed after a period of discontinuous transmission, the predictive quantiser memories in the speech encoder of the uplink mobile station are started from their reset states. To reflect this operation to the other elements of the TFO connection, the following steps are performed. The dequantisation operation in the predictive dequantising block 203 are also started from the reset state. A re-encoding is performed to the decoded speech parameters during the first frame of normal speech to keep the memory 205a of the re-encoding quantiser block 205 of the UTR and the memory of the dequantiser block of the speech decoder of the DMS synchronised, to prevent any audibly annoying effects caused by loss of synchronisation.

For the re-encoding of this first speech frame, the quantiser 205 uses the memory contents left by the the last re-encoded comfort noise frame. After re-encoding, the contents of the quantiser block 205 memory 205a are copied to the memory 203a of the dequantiser 203 for the next frame. In the second and any further good speech frames, the parameters extracted in the extraction block 202 are forwarded to TFO packing and the decoding of speech parameters at the decoding block 203 continue as in normal TFO operation.

Figure 6 illustrates as an example a method according to a further advantageous embodiment of the invention. The figure illustrates a single cycle of processing a set of parameters during tandem free operation, in a BFH/CNI processing situation. First, in step 310, the parameters are received, after which the parameters are decoded in step 320. The decoded parameters are processed in step 330. In this processing step, BFH/CNI processing is performed as described elsewhere in this

specification. The processed parameters are re-encoded in step 340. The state of the encoder is at least in part transferred to the decoder by updating of the decoding block memory in step 350. For further transmission of the received parameters, at least part of them are replaced by processed and re-encoded parameters in step 360, after which the parameters are transmitted further in the transmission path in step 370.

A benefit of this invention is, that it makes possible proper processing of CNI and BFH during tandem free operation, when predictive or more generally non-stateless quantisers are used in the transmitting mobile station. In prior art solutions, the combination of predictive quantisers and BFH/CNI is not possible in tandem free operation without audible and annoying artefacts.

The functional blocks implementing the method according to the invention can be located in many different network elements. The functional blocks can advantageously be located in a so-called transcoder unit (TRCU). The transcoder unit can be a standalone unit, or it can be integrated for example in a base station (BS), in a base station controller (BSC), or in a mobile switching center (MSC). However, the invention is not limited only to implementation in a transcoder unit.

The invention is not limited to such a system, where all speech parameters are encoded by predictive encoders. In a mobile telecommunications system, where only a part of speech parameters are encoded by a predictive encoder and some speech parameters are encoded by stateless encoders, a speech decoder according to an advantageous embodiment of the invention may, for example, process speech parameters encoded by stateless encoders in a way known in prior art, and predictively encoded parameters in an inventive way described previously.

The invention is not limited to the GSM system only. The GSM system is presented only as an example in this specification. The invention can be applied in any digital cellular mobile telecommunication system, such as the so-called third generation cellular systems, which were under development at the time this specification was filed.

In this specification and in the following Claims, the term "non-stateless" denotes a decoder or an encoder having functional states, i.e. being dependent in at least some degree on at least some of the previous inputs, in addition to the most recent or present input. The term "speech coding processing unit" denotes a functional entity,

which decodes encoded speech parameters and/or converts the coding of encoded speech parameters from a first coding method to a second coding method.

In view of the foregoing description it will be evident to a person skilled in the art that various modifications may be made within the scope of the invention. While a preferred embodiment of the invention has been described in detail, it should be apparent that many modifications and variations thereto are possible, all of which fall within the true spirit and scope of the invention.

Claims

1. A method for processing speech signal parameters encoded by a non-stateless encoder, **characterized** in that the method comprises steps of
decoding the encoded speech parameters using a non-stateless decoder,
processing the decoded speech parameters,
re-encoding the processed decoded speech parameters using a second non-stateless encoder,
updating the state of the non-stateless decoder at least in part with the state of the second non-stateless encoder, and
replacing at least one encoded speech parameter with a re-encoded speech parameter to produce processed encoded speech parameters.

2. A method according to claim 1, **characterized** in that before said re-encoding step, the state of said second non-stateless encoder is updated with the state of the non-stateless decoder.

3. A method according to claim 1, **characterized** in that in said speech parameter processing step, comfort noise information is converted into the decoded speech parameters.

4. A method according to claim 1, **characterized** in that in said speech parameter processing step, bad frame handling information is converted into the decoded speech parameters.

5. A speech coding processing unit for decoding encoded speech parameters and producing a decoded time domain speech signal and encoded speech parameters representing the signal, **characterized** in that the unit comprises
a non-stateless decoding block for decoding the encoded speech parameters,
a speech parameter processing block for processing the decoded speech parameters, and
a non-stateless encoding block for encoding the processed speech parameters for producing the encoded speech parameters representing the signal.

6. A speech coding processing unit according to claim 5, **characterized** in that said non-stateless decoding block is a predictive dequantiser, and
said non-stateless encoding block is a predictive quantiser.

7. A speech coding processing unit according to claim 5, **characterized** in that said speech parameter processing block is a comfort noise processing block.

8. A speech coding processing unit according to claim 5, **characterized** in that said speech parameter processing block is a bad frame handling block.

9. A speech coding processing unit according to claim 5, **characterized** in that the unit is a transcoder unit.

10. A speech coding processing unit according to claim 5, **characterized** in that the unit is an uplink transcoder unit.

11. A telecommunications network element for receiving encoded speech parameters and transmitting a time-domain speech signal with embedded encoded speech parameters, **characterized** in that the network element comprises a speech coding processing unit having
a non-stateless decoding block for decoding the encoded speech parameters,
a speech parameter processing block for processing the decoded speech parameters, and
a non-stateless encoding block for encoding the processed speech parameters for producing the the embedded encoded speech parameters.

12. A telecommunications network element according to claim 11, **characterized** in that said non-stateless decoding block is a predictive dequantiser, and
said non-stateless encoding block is a predictive quantiser.

13. A telecommunications network element according to claim 11, **characterized** in that said speech parameter processing block is a comfort noise processing block.

14. A telecommunications network element according to claim 11, **characterized** in that said speech parameter processing block is a bad frame handling block.

15. A telecommunications network element according to claim 11, **characterized** in that said unit is a transcoder unit.

16. A telecommunications network element according to claim 11, **characterized** in that said unit is an uplink transcoder unit.

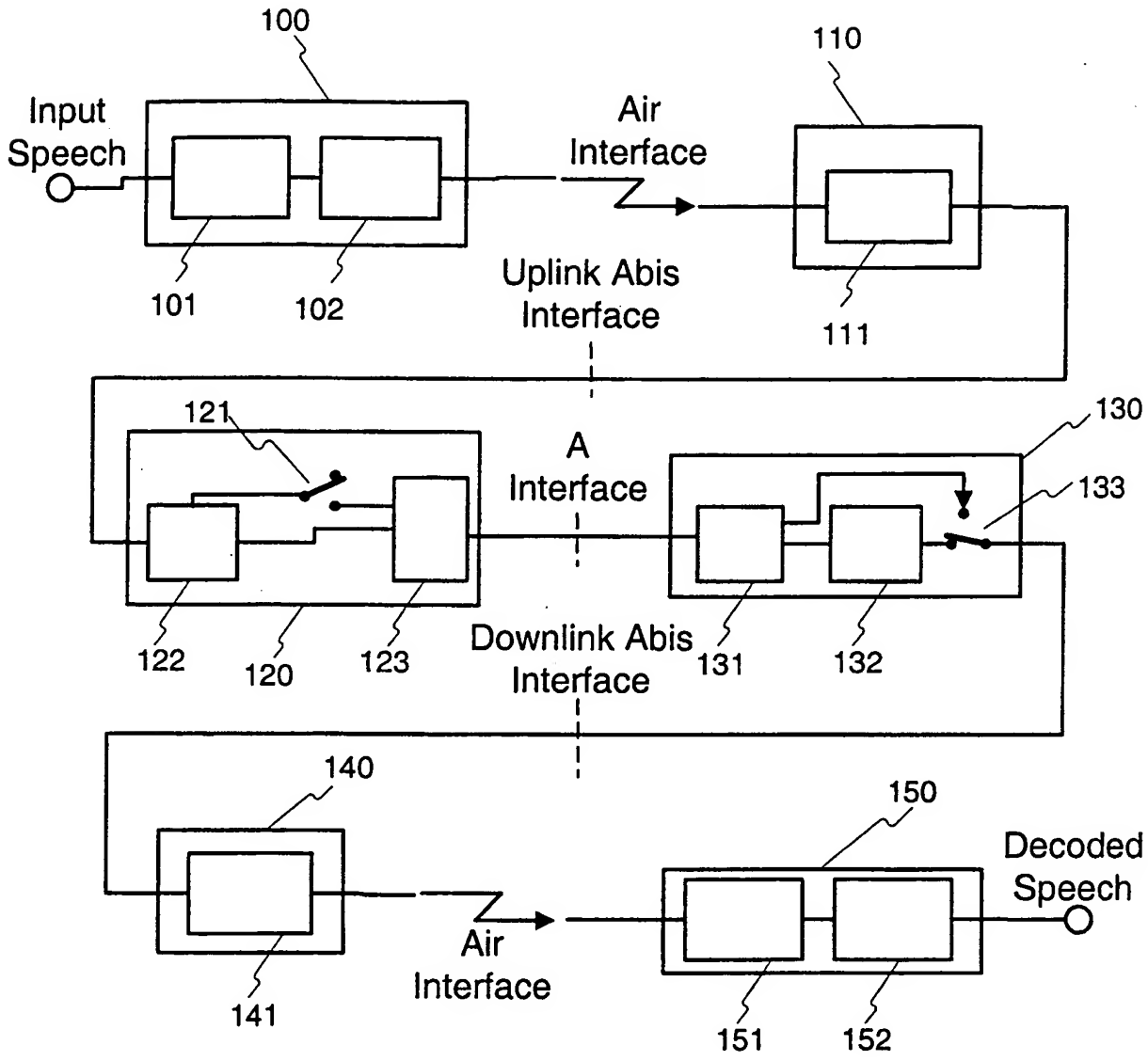


Fig. 1
PRIOR ART

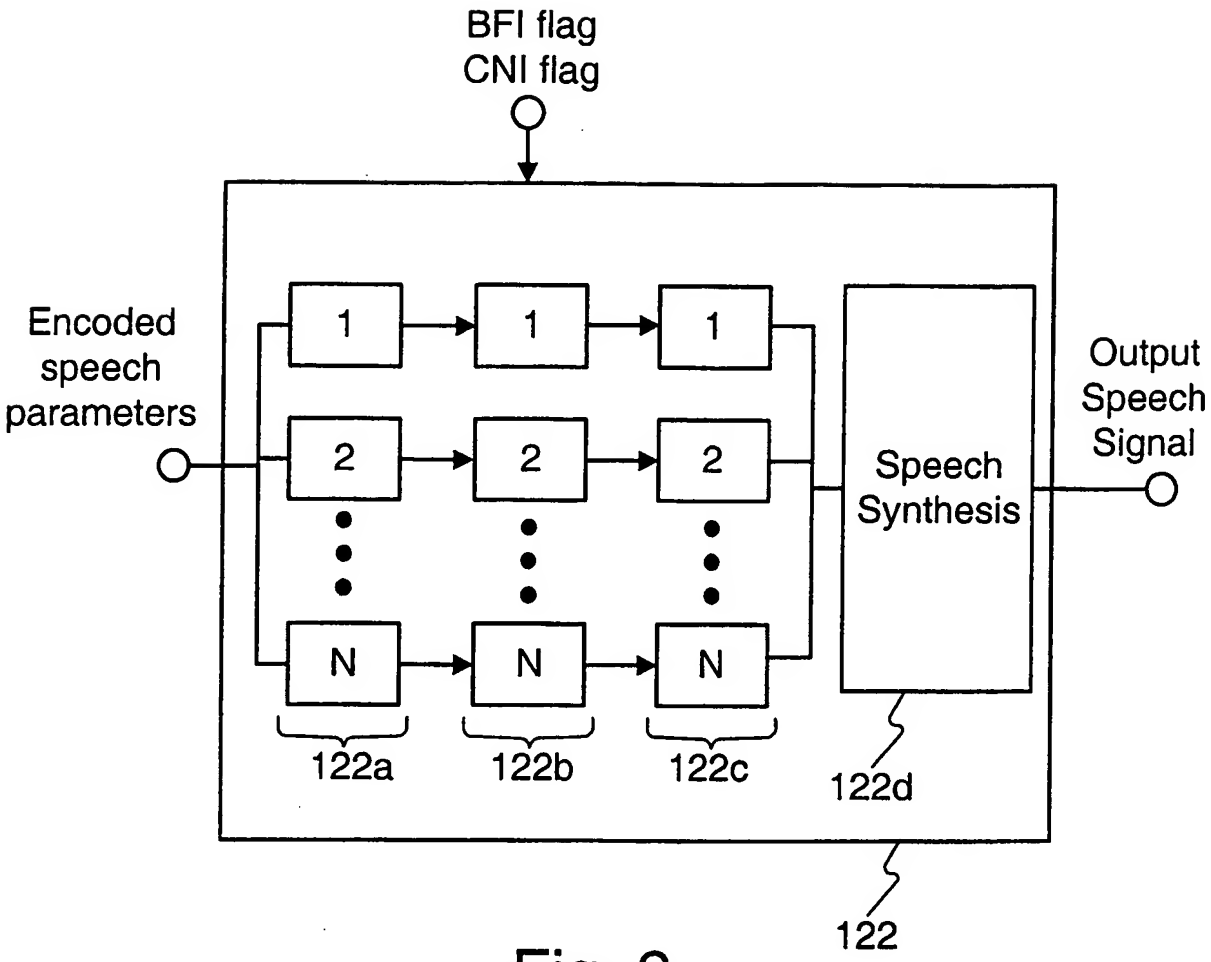


Fig. 2
PRIOR ART

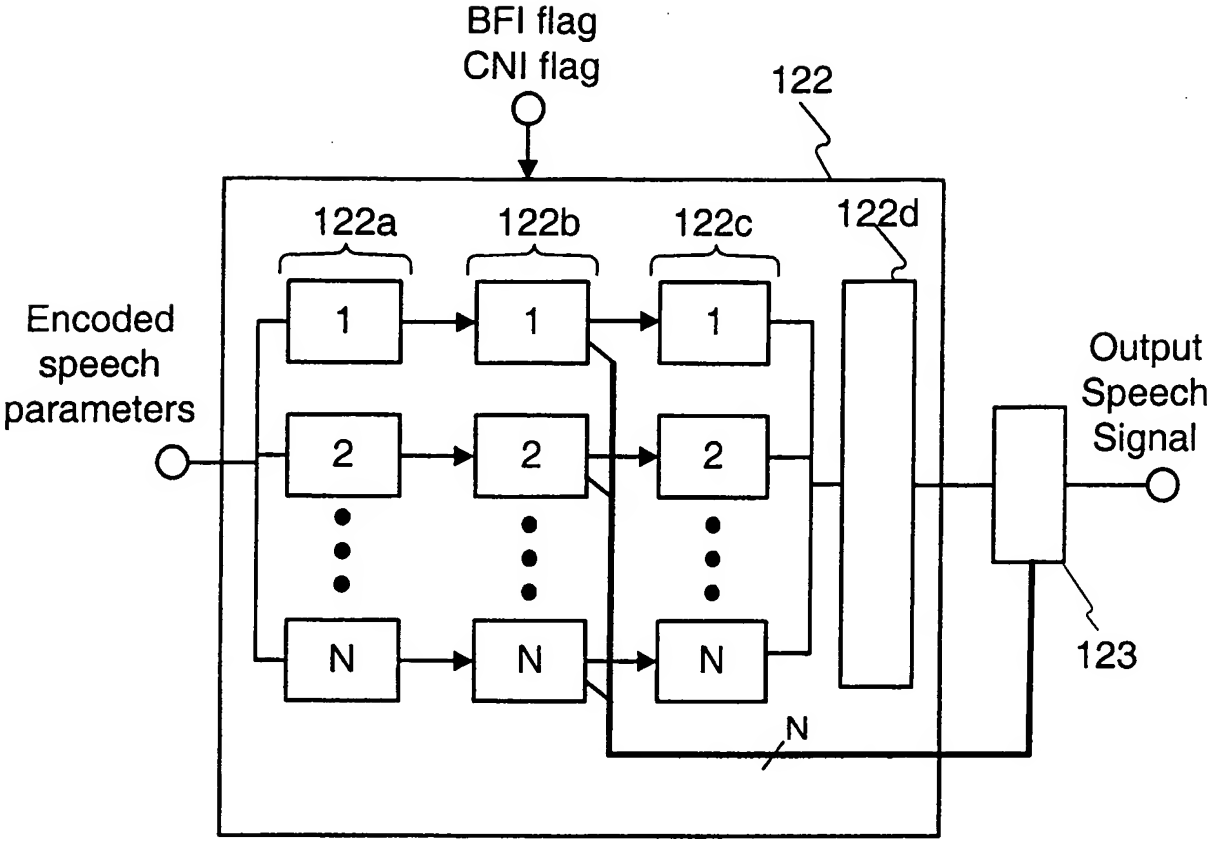


Fig. 3
PRIOR ART

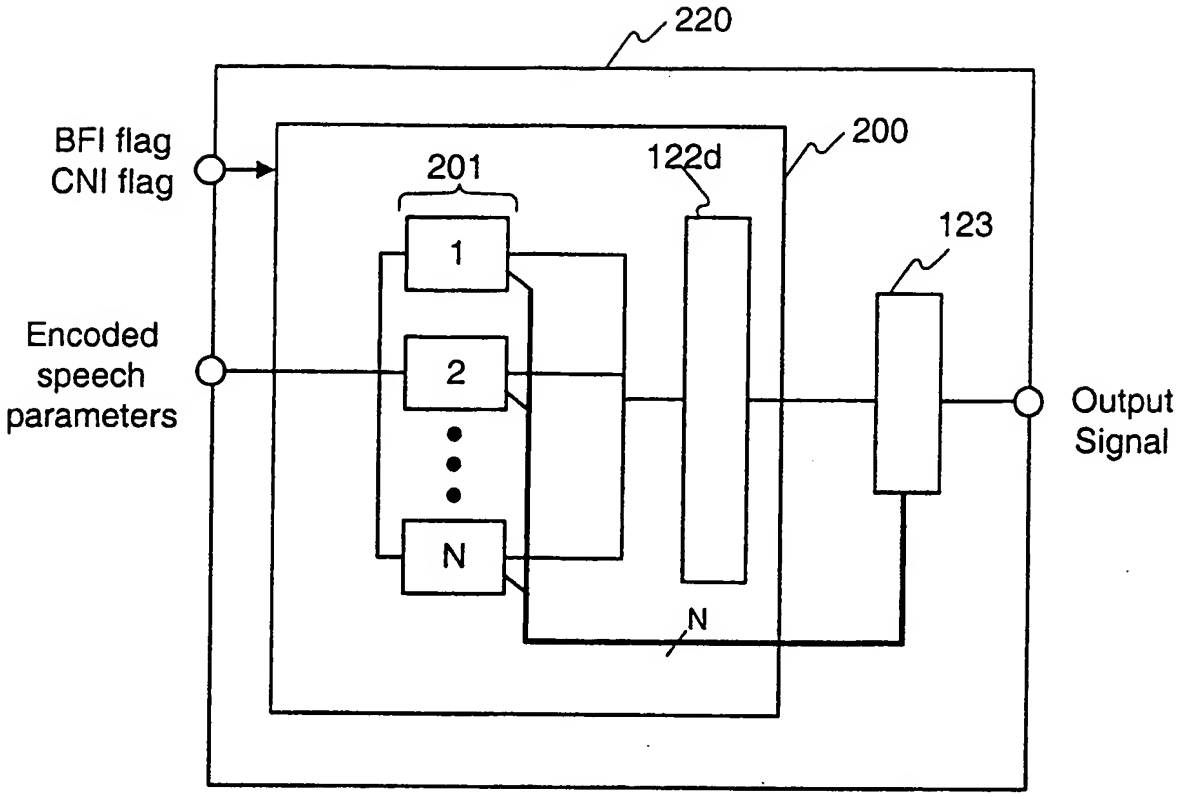

220
200
BFI flag
CNI flag
201
122d
123
1
Encoded
speech
parameters
2
N
N
Output
Signal

Fig. 4

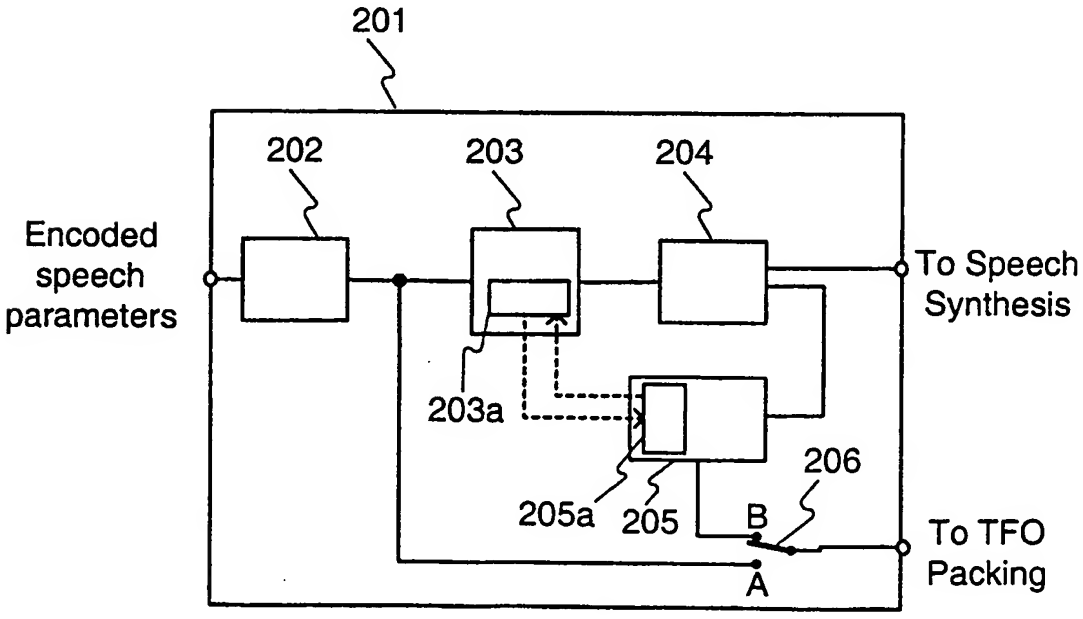

201
202
203
204
Encoded
speech
parameters
To Speech
Synthesis
203a
206
205a
205
B
To TFO
Packing
A

Fig. 5

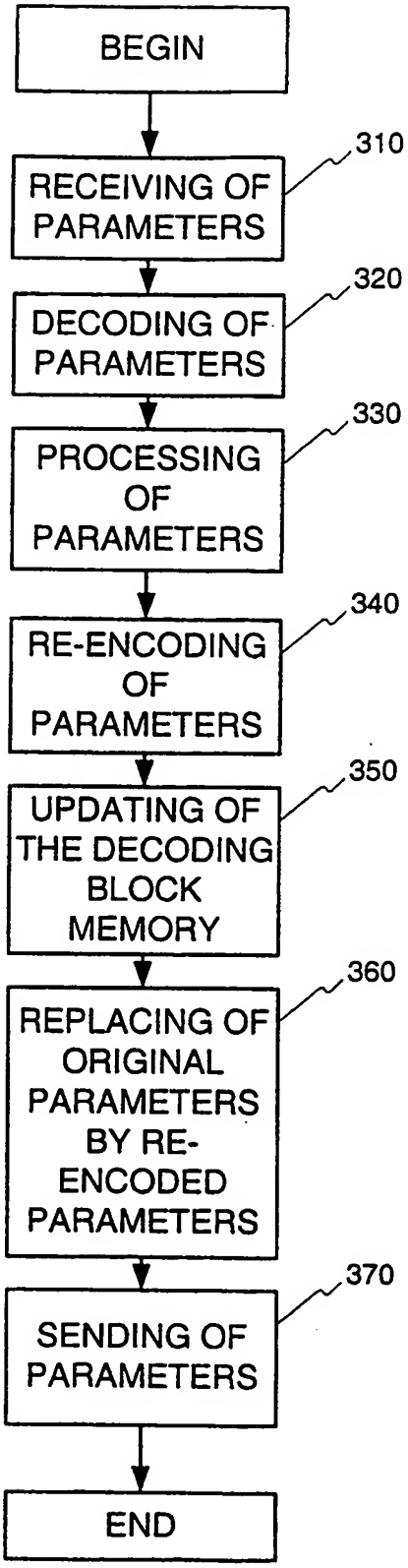

BEGIN
RECEIVING OF PARAMETERS
310
DECODING OF PARAMETERS
320
PROCESSING OF PARAMETERS
330
RE-ENCODING OF PARAMETERS
340
UPDATING OF THE DECODING BLOCK MEMORY
350
REPLACING OF ORIGINAL PARAMETERS BY RE-ENCODED PARAMETERS
360
SENDING OF PARAMETERS
370
END

Fig. 6

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 November 2003 (27.11.2003)

PCT

(10) International Publication Number
WO 03/098598 A1

(51) International Patent Classification[7]: G10L 11/04, 11/00, 21/00, 19/00

(21) International Application Number: PCT/US03/06335

(22) International Filing Date: 26 February 2003 (26.02.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/145,533 13 May 2002 (13.05.2002) US

(71) Applicant: **CONEXANT SYSTEMS, INC.** [US/US]; 4311 Jamboree Road, Newport Beach, CA 92660 (US).

(72) Inventors: **BENYASSINE, Adil**; 40 Dinuba, Irvine, CA 92602 (US). **SHLOMOT, Eyal**; 216 Quincy Avenue, n°1, Long Beach, CA 90803 (US). **SU, Huan-Yu**; 3009 Calle Frontera, San Clemente, CA 92673 (US). **THYSSEN, Jes**; 7 Novilla, Laguna Niguel, CA 93677 (US). **GAO, Yang**; 26586 San Torini Road, Mission Viejo, CA 92692 (US).

(74) Agent: **FARJAMI, Farshad**; Farjami & Farjami LLP, 16148 Sand Canyon, Irvine, CA 92618 (US).

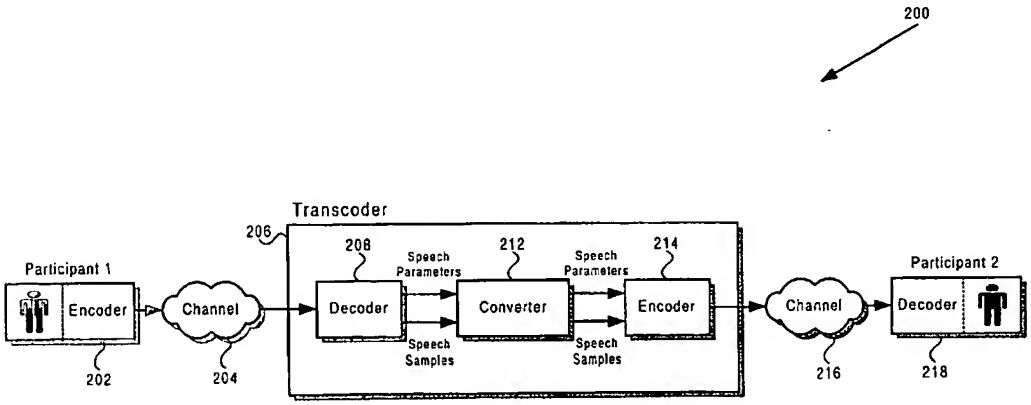
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TRANSCODING OF SPEECH IN A PACKET NETWORK ENVIRONMENT



(57) Abstract: There is provided transcoding of speech in a packet network environment. A decoder configured to receive a first bit-stream encoded according to a first coding scheme. The decoder decodes the bit-stream according to the first coding scheme, generates a plurality of first speech samples, and extracts a plurality of first speech parameters, which may include spectral characteristics, energy, pitch and/or pitch gain. A converter then converts the plurality first speech samples and plurality of first speech parameters to a plurality of second speech samples and a plurality of second speech parameters for use according to a second coding scheme. The first and second coding schemes may be, for example, G.711, G.723.1, G.726 or G.729, and may be parametric or non-parametric. An encoder receives the plurality of second speech samples and plurality of second speech parameters and generates a second bit-stream according to the second coding scheme.

WO 03/098598 A1

TRANSCODING OF SPEECH IN A PACKET NETWORK ENVIRONMENT

RELATED APPLICATIONS

The present application is a continuation-in-part of United States application serial number 09/547,832, filed April 12, 2000, which claims the benefit of provisional United States application serial number 60/128,873, filed April 12, 1999, which are hereby fully incorporated by reference in the present application.

BACKGROUND OF THE INVENTION

1. FIELD OF THE INVENTION

The present invention relates generally to the field of speech coding and, more particularly, to transcoding of speech in a packet network environment.

2. RELATED ART

The explosive growth of the Internet has been accompanied by a growing interest in using this traditionally data-oriented network for voice communication in accordance with voice-over-packet ("VoP"). The packetizing of voice signals for transmission over a packet network has been recognized as a less expensive, yet effective, alternative to traditional telephone service. The term VoP is an umbrella term that can include, for example, VoIP and other types of services utilizing packetized voice data.

One challenge facing the expansion of VoP is the need to connect diverse types of networks with greater effectiveness. More specifically, because different networks may be using different standards to encode, compress and packetize speech, a transcoding procedure has to be performed in order for a meaningful connection between networks to be achieved. Typically, voice data encoded according to one standard from a transmitting participant communicating in one network has to be converted to the standard used by the receiving participant communicating under the guidelines of another network. For example, a transmitting participant's speech may be encoded according to G.723.1 specifications while the receiving participant uses G.729. In order for the data from the transmitting participant to be understood by the receiving participant, the bit-stream from the transmitting participant has to be converted from G.723.1 format to G.729 format.

In conventional transcoding approaches, encoded data from the transmitting participant is decoded according to the coding method used by the transmitting participant. The decoded data is then re-encoded in accordance with the coding method used by the receiving participant. In the re-encoded form, the data is transmitted to the receiving participant. Known transcoding schemes, however, suffer numerous serious inadequacies. For example, the decoding and re-encoding of the speech signal (a "tandem" process), reduces the quality of the speech. More particularly, the tandem operation of the post-filter, common in low bit-rate speech decoders, can generate objectionable spectral distortion and degrade the speech quality significantly.

Another drawback of known transcoding schemes is the undesirable delay resulting from the re-encoding step. Typically, re-encoding of the decoded bit-stream requires that the speech signal characteristics be evaluated. As such, parameters including energy, spectral characteristics and pitch,

for example, have to be extracted from the bit-stream and used to re-encode the signal. Furthermore, in addition to delay, the need to extract these parameters as part of the re-encoding step introduces greater complexity to the system.

Thus, there is an intense need in the art for a transcoding method, and related system, which can overcome the shortcomings of known transcoding schemes and provide for more effective means by which transcoding between networks can be achieved.

SUMMARY OF THE INVENTION

In accordance with the purpose of the present invention as broadly described herein, there is provided transcoding of speech in a packet network environment. In one exemplary aspect of the present invention, a speech transcoder capable of transcoding a first bit-stream generated from a speech signal is disclosed. The transcoder includes a decoder configured to receive the first bit-stream, which has been encoded based on a first coding scheme. For example, the speech signal may have been encoded according to G.711, G.723.1, G726 or G.729, and may be parametric or non-parametric. The decoder extracts a plurality of first speech parameters from the first bit-stream, which may include, for example, parameters relating to spectral characteristics, energy, pitch and/or pitch gain of the speech signal. The decoder also decodes the first bit-stream according to the first coding scheme and generates a plurality of first speech samples. In certain configurations, the decoder may include a post-filter element, which may be disabled to reduce system complexity and to improve the speech-quality of a speech signal generated by a subsequent re-encoding process.

The plurality of first speech samples and the plurality of first speech parameters are then transmitted to a converter capable of converting the plurality first speech samples and plurality of first speech parameters to a plurality of second speech samples and a plurality of second speech parameters for use according to a second coding scheme. The second coding scheme may be G.711, G.723.1, G726 or G.729, for example, and may be parametric or non-parametric. Following conversion by the converter, the plurality of second speech samples and the plurality of second speech parameters are transmitted to an encoder. The encoder receives the plurality of second speech samples and plurality of speech parameters and generates a second bit-stream, the second bit-stream being encoded based on the second coding scheme. In certain configurations, the encoder may include a noise suppressor element, which may be disabled to reduce system complexity and to improve the speech-quality of a speech signal. It is appreciated that by extracting the speech parameters from the first bit-stream, converting the speech parameters, and providing the converted speech parameters to the encoder avoids a re-evaluation of speech parameters during the encoding process, achieving many advantageous results, such as reduced system complexity and less delay.

These and other aspects of the present invention will become apparent with further reference to the drawings and specification, which follow. It is intended that all such additional systems, methods, features and advantages be included within this description, be within the scope of the present invention, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

The features and advantages of the present invention will become more readily apparent to those ordinarily skilled in the art after reviewing the following detailed description and accompanying drawings, wherein:

FIG. 1 illustrates a block diagram of a packet-based network in which various aspects of the present invention may be implemented;

FIG. 2 illustrates a block diagram of a transcoding system in accordance with one embodiment;

FIG. 3 illustrates a block diagram of a conference bridge utilizing a transcoding system in accordance with one embodiment;

FIG. 4 illustrates a block diagram of a component of a conference bridge utilizing a transcoding system in accordance with one embodiment; and

FIG. 5 illustrates an exemplary flow diagram of a transcoding method utilizing the transcoding system of FIG. 2.

DESCRIPTION OF EXEMPLARY EMBODIMENTS

The present invention may be described herein in terms of functional block components and various processing steps. It should be appreciated that such functional blocks may be realized by any number of hardware components and/or software components configured to perform the specified functions. For example, the present invention may employ various integrated circuit components, e.g., memory elements, digital signal processing elements, logic elements, and the like, which may carry out a variety of functions under the control of one or more microprocessors or other control devices. Further, it should be noted that the present invention may employ any number of conventional techniques for data transmission, signaling, signal processing and conditioning, tone generation and detection and the like. Such general techniques that may be known to those skilled in the art are not described in detail herein.

It should be appreciated that the particular implementations shown and described herein are merely exemplary and are not intended to limit the scope of the present invention in any way. Indeed, for the sake of brevity, conventional data transmission, signaling and signal processing and other functional and technical aspects of the communication system (and components of the individual operating components of the system) may not be described in detail herein. Furthermore, the connecting lines shown in the various figures contained herein are intended to represent exemplary functional relationships and/or physical couplings between the various elements. It should be noted that many alternative or additional functional relationships or physical connections may be present in a practical communication system.

FIG. 1 depicts an exemplary communication environment 100 that is capable of supporting the transmission of packetized voice information. A packet network 102, e.g., a network conforming to the Internet Protocol ("IP"), may support Internet telephony applications that enable a number of participants to conduct voice calls in accordance with conventional voice-over-packet techniques. In a practical environment 100, packet network 102 may communicate with conventional telephone networks, local area networks, wide area networks, public branch exchanges, and/or home networks in a manner that enables participation by users that may have different communication devices and different communication service providers. For example, in FIG. 1, Participant 1 and Participant 2 communicate with packet network 102 (either directly or indirectly) via the transmission of packets that contain voice data. Participant 3 communicates with packet network 102 via a gateway 104, while Participant 4 communicates with packet network 102 via a gateway 106.

In the context of this description, a gateway is a functional element that converts voice data into packet data. Thus, a gateway may be considered to be a conversion element that converts conventional voice information into a packetized form that can be transmitted over a packet network. A gateway may be implemented in a central office, in a peripheral device (such as a telephone), in a local switch (e.g., one associated with a public branch exchange), or the like. The functionality and operation of such gateways are well known to those skilled in the art and will therefore not be described in detail. It will be appreciated that the present invention can be implemented in conjunction with a variety of conventional gateway designs.

Environment 100 may include any number of transcoders that enable communication between participants using different speech coding standards. For example, a transcoder 108 may be included in packet network 102. Transcoder 108 may be implemented in a central office or

maintained by an Internet service provider ("ISP"). In this manner, the voice data from a number of packet-based participants, e.g., Participants 1 and 2, can be processed by transcoder 108 without having to perform the conversions normally performed by gateways.

As another example, a transcoder 110 may be associated with or included in a gateway, e.g., gateway 104. In this configuration, transcoder 110 may be capable of receiving and processing voice-over-packet data and conventional voice signals. Eventually, gateway 104 enables Participant 3, through transcoder 110, to communicate with packet network 102 and a participant coupled to packet network 102, e.g., Participant 1 or 2.

In accordance with the present invention, a packet-based transcoder may be deployed in a telephony system to facilitate communication between participants using different standards or techniques of speech coding. As is known, a given packet-based voice channel, for example, may employ one of a number of different speech coding/compression standards. Various speech coding standards are generally known to those skilled in the art and may include, for example, G.711, G.726, G.728, G.729(A), G.723.1, Global System for Mobile Communication ("GSM"), selectable mode vocoder ("SMV"), and adaptive multi rate ("AMR") coding, the specifications for which are hereby incorporated by reference.

The particular standard utilized for a given call may depend on the participant's Internet service provider, telephone service provider, design of the participant's peripheral device, and other factors. Consequently, a practical transcoder, such as transcoder 108 or 110, may be capable of handling speech that has been encoded by various standards. In addition, such a transcoder should be capable of handling speech that has not been encoded.

FIG. 2 illustrates exemplary communication system 200 for transcoding in accordance with one embodiment of the present invention. As shown in communication system 200, a first participant (i.e., Participant 1) is communicating with a second participant (i.e., Participant 2) through transcoder 206. Participant 1 is coupled to transcoder 206 via channel 204, and Participant 2 is coupled to transcoder 206 via channel 216.

In the illustrated embodiment, voice data from Participant 1 may be encoded by encoder 202 and sent to transcoder 206 via channel 204. As discussed above, depending on such factors as the participant's Internet service or telephony service, for example, the voice data from participant 1 may need to be compressed and encoded by encoder 202 using a suitable coding standard. For example, channel 204 may be an Internet-based packet network, in which case encoder 202 may use a suitable packet format to packetize the voice data. In such case, the output data from encoder 202 transmitted over channel 204 will include encoded digital data, in the form of a bit-stream, in accordance with one or more encoding standards, e.g., G.723.1 or G.729. Alternatively, channel 204 may function as a local link, coupling Participant 1 to transcoder 206, in which case encoder 202 may digitize the voice data from Participant 1 without encoding, and digitized data is transmitted over channel 204.

The bit-stream from Participant 1 arriving at transcoder 206 via channel 204 is initially inputted into, and processed by decoder 208, which is configured to decode the bit-stream according to the coding method for the transmitting participant, i.e., Participant 1. Thus, if the voice data from Participant 1 was encoded by encoder 202 using G.723.1, for instance, then decoder 208 would decode the bit-stream accordingly. In one embodiment, post-filter element of decoder 208 (not shown) may be disabled or its capabilities reduced to minimize the degradation frequently found with

conventional decoding algorithms utilizing post-filtering.

In addition to generating speech samples from the bit-stream (i.e., the decoded bit-stream), decoder 208 is also configured to extract certain speech parameters from the bit-stream. The speech parameters, which are also referred to as "side information" in the present application, may include, for example, the energy, the spectral characteristics, the pitch and pitch gain of the speech signal. Thereafter, in addition to the speech samples, the speech parameters (or the side information) are transmitted by decoder 208 to converter 212.

Continuing with FIG. 2, the speech samples and speech parameters inputted into converter 212 are suitably processed and converted for eventual encoding by an encoder according to the standard suitable for the receiving participant. The conversion performed by converter 212 may be based on the speech samples and/or at least one of the parameters, for example, received from decoder 208. As part of the conversion process, the speech samples may be modified into a format suitable for re-encoding by encoder 214. For example, in instances where Participants 1 and 2 are using coding standards having different frame structures, converter 212 may resize the frames, to provide the speech samples according to a proper frame size for use by encoder 214. Following conversion by converter 212, the speech information, comprising the converted speech samples and speech parameters, is transmitted to encoder 214. It should be noted that in some embodiments, decoder 208 may only provide the speech samples to converter 212 and no speech parameters (or side information). For example, when the speech signal is coded according to a non-parametric coding scheme, such as G.711, G.726, G.728, etc., converter 212 receives the speech samples from decoder 208 and converts the speech samples to provide them according to a proper frame size for use by encoder 214.

Encoder 214 is configured to encode the speech information according to the standard used by the receiving participant, i.e., Participant 2 in the present example. Thus, if Participant 2 utilizes a selectable mode vocoder ("SMV"), for example, then encoder 214 would encode the bit-stream according to the SMV standard. According to the present invention, encoder 214 can be configured to encode speech information using the speech parameters extracted by decoder 208 and processed by converter 212. In this manner, parameters such as the energy, the spectral characteristics, the pitch and pitch gain of the speech signal, which are conventionally needed to re-encode the speech information by encoder 214, do not have to be re-extracted from the speech samples by encoder 214. Thus, encoder 214 does not have to perform such parameter estimation tasks as spectral analysis, pitch analysis, and the like, or encoder 214 may only have to perform lower complexity parameter estimation tasks. As a result, the transcoding scheme of various embodiments of the present invention substantially reduces the processing power, minimizes delay, and reduces overall system complexity when compared to conventional transcoding schemes. In one embodiment, the noise suppression capability of encoder 214 may be disabled in order to further reduce the system's complexity. Additionally, because the speech parameters are extracted during the initial decoding step for use during the re-encoding step, degradation of the signal resulting from, for example, spectral and pitch re-evaluation is avoided. Following coding by encoder 214, the bit-stream is transmitted to the receiving participant, i.e., Participant 2, via channel 216 in the format suitable for use by decoder 218, which then decodes the bit-stream.

Referring now to FIG. 3, exemplary communication system 300 is used to illustrate a

conference bridge using transcoding techniques of the present invention, in accordance with one embodiment. More particularly, communication system 300 shows how the present invention can be used to transcode and mix speech signals from two or more transmitting participants to a receiving participant, where each transmitting participant may be using a different coding scheme from the other. In communication system 300, Participants 1, 2 and 3 are coupled to conference bridge 306 via channels 304, 316 and 322, respectively. It is appreciated that in the present example, Participants 1 and 3 are both communicating with Participant 2 at the same time.

Continuing with FIG. 3, speech from participant 1 is encoded by encoder 302 into a format suitable for transmission over channel 304 to decoder 308. Similarly, encoder 320 encodes speech from Participant 3 into a format suitable for transmission over channel 322 to decoder 324. Both decoders 308 and 324 can be configured to decode the incoming bit-streams, such as those coming from Participants 1 and 3, according to the coding schemes used by the transmitting participants and to generate speech samples from the bit-streams. Decoders 308 and 324 may also extract speech parameters, from the bit-stream, or generate the speech parameters if the speech was originally encoded according to a non-parametric standard.

Following decoding, the speech samples and the speech parameters for both Participants 1 and 3 are inputted into converter/mixer 312. Converter/mixer 312 can be configured to convert, combine and mix the inputted speech samples and the speech parameters to generate a single speech information suitable for encoding according to the coding scheme used by the receiving participant, i.e., Participant 2.

Depending on the various coding methods used by the transmitting participants, converter/mixer 312 may need to take into account frame size and other factors in order to generate a bit-stream suitable for encoding by the receiving participant. For example, G.723.1 uses a frame size of 30 ms, and G.729 uses a frame size of 10 ms. Thus, a common frame structure may be established to enable effective mixing of the speech samples from decoders 308 and 324. For example, if at least one of the input channels is encoded using G.723.1, then a 30 ms frame may be established. Alternatively, a frame size equal to the least common multiple might be used. In a case where one channel is encoded using G.723.1 (30 ms frame), for example, and another encoded using G.4k (20 ms frame), a 60 ms frame may be established. Once a frame size is determined, the speech samples and the speech parameters can be properly interpolated and aligned during mixing.

Application serial number 09/547,832, filed April 12, 2000, which is incorporated by reference into the present application, discloses methods by which speech parameters are mixed and interpolated are known and may be used by converter/mixer 312 to mix the speech parameters inputted from decoders 308 and 324. For example, the spectrums of two signals may be summed using a weighted addition. A similar method may be used to mix other parameters, such as pitch and energy.

Once converter/mixer 312 has converted and mixed the signal from decoder 308 with the signal from decoder 324 to generate a combined bit-stream, the bit-stream is transmitted to encoder 314. Converter/mixer 312 can also provide encoder 314 with the speech parameters extracted from the inputted speech signals. Encoder 314 can be configured to re-encode the bit-stream according to the same coding standard used by Participant 2. For example, if Participant 2 uses G.726, then encoder 314 would re-encode the speech information according to G.726. Encoder 314 may use the

parameters extracted by decoders 308 and 324 in order to re-encode the speech information, thus bypassing the need for spectral and pitch re-evaluation during the re-encoding process. In this manner, the complexity, processing demands, and time delay associated with such re-evaluation steps are avoided. Following re-encoding by encoder 314, the speech signal is transmitted via channel 316 to Participant 2, where decoder 318 decodes the signal.

Referring now to FIG. 4, exemplary communication system 400 is used to illustrate a component of a conference bridge using transcoding techniques of the present invention, in accordance with one embodiment. More particularly, communication system 400 shows how the present invention provides an effective means for transcoding inputted speech signals having been encoded according to a non-parametric coding standard, such as G.711, G.726, and G.728, for example. As shown in FIG. 4, communication system 400 includes channel 404, conference bridge 406 and channel 416. It is appreciated that channels 404 and 416 are respectively equivalent to channels 204 and 216 of communication system 200 illustrated in FIG. 2.

As shown, a speech signal transmitted to conference bridge 406 via channel 404 is decoded by decoder 408 to generate speech samples from the incoming bit-stream. Decoder 408 may also extract speech parameters from the bit-stream to generate the speech parameters in instances where the speech was encoded originally using a parametric standard, such as G.729 or G723.1. However, it is appreciated that non-parametric speech coding standards, for example G.711, G.726 and G.728, typically do not quantize various speech-related parameters, such as the signal pitch and spectrum. As a result, these parameters may not be extracted by decoder 408 directly from the bit-stream during the decoding process. In such instances, as shown in FIG. 4, the speech samples may be diverted to parameter extraction module 410, which extracts the desired speech-related parameters (or the side information) for subsequent use by encoder 414, as described below. Thus, parameter extraction module 410 can be configured to extract data regarding the signal energy, spectral characteristics, pitch and pitch gain, and the like, and to provide such parameters to converter/mixer 412.

The decoded speech samples from decoder 408 and the speech parameters from either decoder 408 or parameter extraction module 410 are inputted into converter/mixer 412. As shown in FIG. 4, converter/mixer 412 also receives speech samples and the speech parameters (or side information) 420 from other decoding devices (not shown). Converter/mixer 412 can be configured to combine and mix the speech samples and the speech parameters from decoder 408 and parameter extraction module 410 with speech samples and speech parameters 420 into a combined bit-stream suitable for use by encoder 414 in the re-encoding process. For example, in order to combine and mix the signals, converter/mixer may resize the frames of the speech samples in order to establish a common frame structure suitable for encoder 414. Converter/mixer 412 can also provide encoder 414 with the speech parameters (or side information) for use in re-encoding the bit-stream.

The combined speech samples and extracted parameters provided by converter/mixer 412 can be used by encoder 414 to re-encode the speech signal according to the coding standard used by the receiving participant (not shown). Accordingly, by using the speech parameters (or side information) provided by converter/mixer 412, encoder 414 bypasses the need for spectral and pitch re-evaluation during the re-encoding process. In this manner, the complexity, processing demands, and time delay associated with such re-evaluation steps are avoided. Following the coding step, the encoded signal is transmitted to the receiving participant via channel 416.

Reference is now made to FIG. 5, which illustrates exemplary transcoding method 500 in accordance with one embodiment. It is appreciated that transcoding method 500 can be performed by a transcoder such as transcoder 206 in FIG. 1, for example. As shown, transcoding method 500 begins at step 510 and continues to step 512 where the bit-stream from a first participant is received.

Following, at step 514, a parameter set is extracted from the bit-stream. For example, the parameter set may include the signal energy, spectral characteristics, pitch and pitch gain, and the like. Next, at step 516, the bit-stream is decoded according to the coding scheme used by the first participant and speech samples are generated. For example, the received bit-stream may be encoded according to G.723.1, in which case the bit-stream is decoded at step 516 according to G.723.1

After the speech samples have been generated at step 516, transcoding method 500 proceeds to step 518 where the speech samples and parameter set are converted into a suitable form for re-encoding. The form to which the speech samples and parameter set are converted may depend on the particular coding scheme used by the receiving participant. At step 520, the converted speech samples are re-encoded in accordance with the coding scheme used by the receiving participant, i.e., the second participant in the present example. As such, if the second participant in the present description uses G.729, for example, then the re-encoding performed at step 520 would be done according to G.729. The re-encoding performed at step 520 can utilize the parameter set extracted from the bit-stream at step 516. Therefore, at step 520, re-encoding can be effectively achieved without having to perform, for example, spectral and pitch re-evaluation, since the information is already available. In this manner, transcoding method 500 provides a number of advantages over conventional transcoding approaches, including lower processing needs, minimal delay, and a reduction in overall system complexity.

The methods and systems presented above may reside in software, hardware, or firmware on the device, which can be implemented on a microprocessor, digital signal processor, application specific IC, or field programmable gate array ("FPGA"), or any combination thereof, without departing from the spirit of the invention. Furthermore, the present invention may be embodied in other specific forms without departing from its spirit or essential characteristics. The described embodiments are to be considered in all respects only as illustrative and not restrictive.

CLAIMS

What is claimed is:

1. A speech transcoder capable of transcoding a first bit-stream generated from a speech signal, said speech transcoder comprising:

a decoder configured to receive said first bit-stream encoded based on a first coding scheme, wherein said decoder extracts a first plurality speech parameters from said first bit-stream, and wherein said decoder decodes said first bit-stream according to said first coding scheme and generates a plurality of first speech samples;

a converter configured to receive said plurality of first speech samples and said plurality of first speech parameters, wherein said converter converts said plurality of first speech samples to a plurality of second speech samples and converts said plurality of first speech parameters to a plurality of second speech parameters for use according to a second coding scheme; and

an encoder configured to receive said plurality of second speech samples and said plurality of second speech parameters, wherein said encoder generates a second bit-stream encoded based on said second coding scheme.

2. The transcoder of claim 1, wherein said converter converts a first frame size of said plurality of first speech samples to a second frame size, wherein said encoder uses said second frame size to generate said second bit-stream according to said second coding scheme.

3. The transcoder of claim 1, wherein said converter transmits said plurality of second speech parameters to said encoder to avoid a re-evaluation of parameters by said encoder and thereby to reduce delay.

4. The transcoder of claim 1, wherein said decoder includes a post-filter element, and wherein said post-filter element is disabled.

5. The transcoder of claim 1, wherein said encoder includes a noise suppressor, and wherein said noise suppressor is disabled.

6. The transcoder of claim 1, wherein said plurality of second speech parameters includes at least one parameter relating to an energy of said speech signal.

7. The transcoder of claim 1, wherein said plurality of first speech parameters includes at least one parameter relating to spectral characteristics of said speech signal.

8. The transcoder of claim 1, wherein said plurality of first speech parameters includes at least one parameter relating to a pitch of said speech signal.

9. The transcoder of claim 1, wherein said plurality of first speech parameters includes at least one parameter relating to a pitch gain of said speech signal.

10. The transcoder of claim 1, wherein said converter transmits said plurality of second speech parameters to said encoder to avoid a re-evaluation of parameters by said encoder and thereby reduce degradation of a speech signal generated from said second bit-stream.

11. A method for transcoding a first bit-stream generated from a speech signal, said speech method comprising:

extracting a plurality of first speech parameters from said first bit-stream;

decoding said first bit-stream according to a first coding scheme to generate a plurality of first speech samples;

converting said plurality of first speech samples to a plurality of second speech samples for use according to a second coding scheme;

converting said plurality of first speech parameters to a plurality of second speech parameters for use according to a second coding scheme; and

encoding said plurality of second speech samples based on said plurality of second speech parameters to generate a second bit-stream encoded based on said second coding scheme.

12. The method of claim 11 further comprising: converting a first frame size of said plurality of first speech samples to a second frame size for use according to said second coding scheme.

13. The method of claim 11, wherein said converting said plurality of first speech parameters to said plurality of second speech parameters is performed to avoid a re-evaluation of parameters during said encoding to reduce delay and complexity.

14. The method of claim 11 further comprising: disabling post-filtering during said decoding.

15. The method of claim 11 further comprising: disabling noise suppression during said encoding.

16. The method of claim 11, wherein said plurality of second speech parameters includes at least one parameter relating to an energy of said speech signal.

17. The method of claim 11, wherein said plurality of first speech parameters includes at least one parameter relating to spectral characteristics of said speech signal.

18. The method of claim 11, wherein said plurality of first speech parameters includes at least one parameter relating to a pitch of said speech signal.

19. The method of claim 11, wherein said plurality of first speech parameters includes at least one parameter relating to a pitch gain of said speech signal.

20. The method of claim 11, wherein said converting said plurality of first speech parameters to said plurality of second speech parameters is performed to avoid a re-evaluation of parameters during said encoding and thereby reduce degradation of a speech signal generated from said second bit-stream.

21. A speech transcoder capable of transcoding a first bit-stream generated from a speech signal, said speech transcoder comprising:

a decoder configured to receive said first bit-stream encoded based on a first coding scheme, wherein said decoder decodes said first bit-stream according to said first coding scheme and generates a plurality of first speech samples;

a parameter extractor module configured to receive said plurality of first speech samples, wherein said parameter extractor module extracts a first plurality speech parameters from said plurality of first speech samples;

a converter/mixer configured to receive said plurality of first speech samples and said plurality of first speech parameters, wherein said converter converts and mixes said plurality of first speech samples to generate a plurality of second speech samples and converts and mixes said plurality of first speech parameters to generate a plurality of second speech parameters for use according to a second coding scheme; and

an encoder configured to receive said plurality of second speech samples and said plurality of second speech parameters, wherein said encoder generates a second bit-stream encoded based on said second coding scheme.

22. The transcoder of claim 21, wherein said converter transmits said plurality of second speech parameters to said encoder to avoid a re-evaluation of parameters by said encoder and thereby to reduce delay.

23. The transcoder of claim 21, wherein said decoder includes a post-filter element, and wherein said post-filter element is disabled.

24. The transcoder of claim 21, wherein said encoder includes a noise suppressor, and said noise suppressor is disabled.

25. The transcoder of claim 21, wherein said plurality of second speech parameters includes at least one parameter relating to an energy of said speech signal.

26. The transcoder of claim 21, wherein said plurality of first speech parameters includes at least one parameter relating to spectral characteristics of said speech signal.

27. The transcoder of claim 21, wherein said plurality of first speech parameters includes at least one parameter relating to a pitch of said speech signal.

28. The transcoder of claim 21, wherein said plurality of first speech parameters includes at least one parameter relating to a pitch gain of said speech signal.

29. The transcoder of claim 21, wherein said converter transmits said plurality of second speech parameters to said encoder to avoid a re-evaluation of parameters by said encoder and thereby reduce degradation of a speech signal generated from said second bit-stream.

30. A speech transcoder capable of transcoding a first bit-stream generated from a speech signal, said speech transcoder comprising:

a decoder configured to receive said first bit-stream encoded based on a first coding scheme, wherein said decoder decodes said first bit-stream according to said first coding scheme and generates a plurality of first speech samples from said bit-stream;

a converter configured to receive said plurality of first speech samples, wherein said converter converts said plurality of first speech samples to a plurality of second speech samples for use according to a second coding scheme; and

an encoder configured to receive said plurality of second speech samples, wherein said encoder generates a second bit-stream encoded based on said second coding scheme.

31. The transcoder of claim 30, wherein said converter converts a first frame size of said plurality of first speech samples to a second frame size, wherein said encoder uses said second frame size to generate said second bit-stream according to said second coding scheme.

FIG. 1

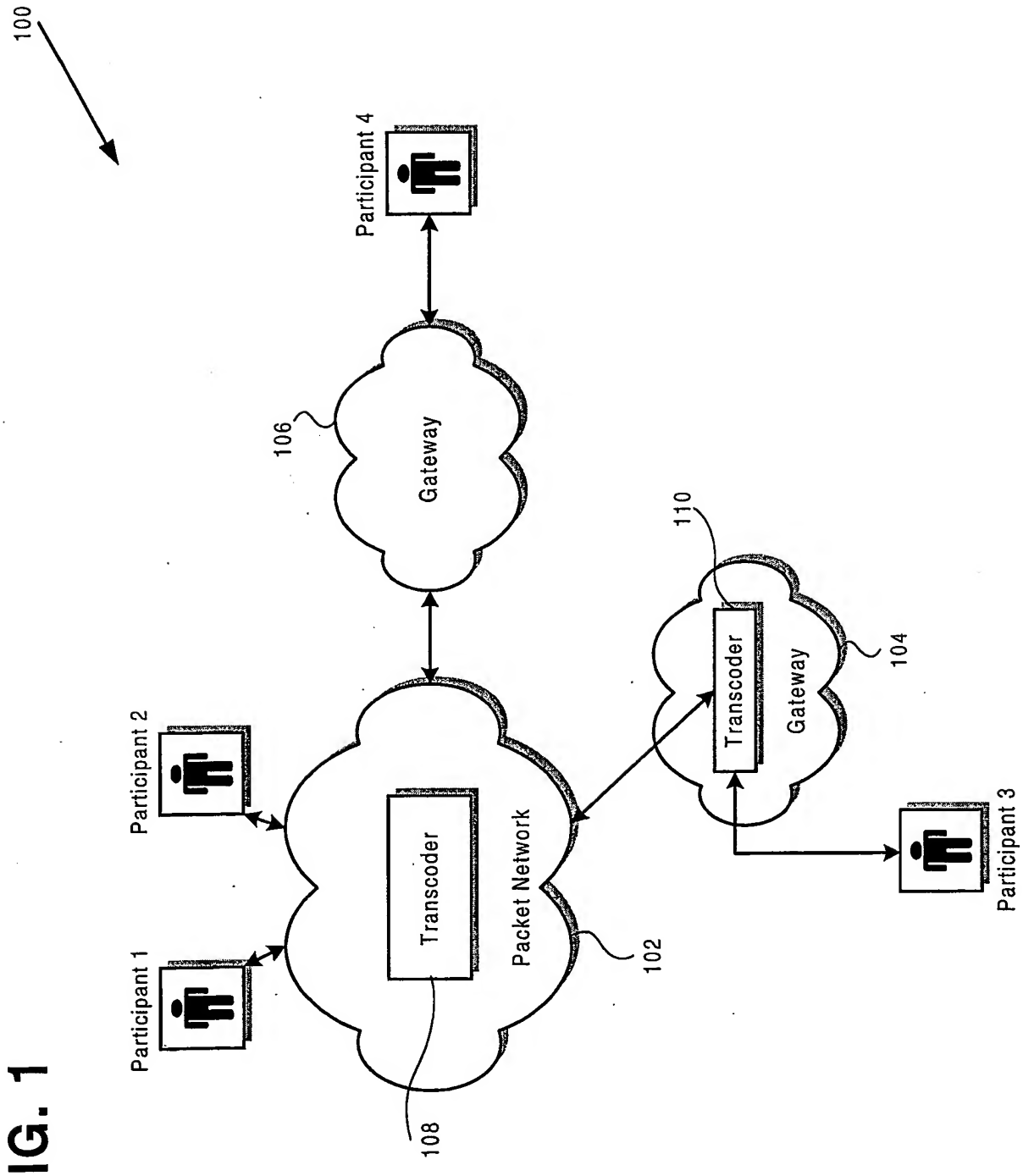
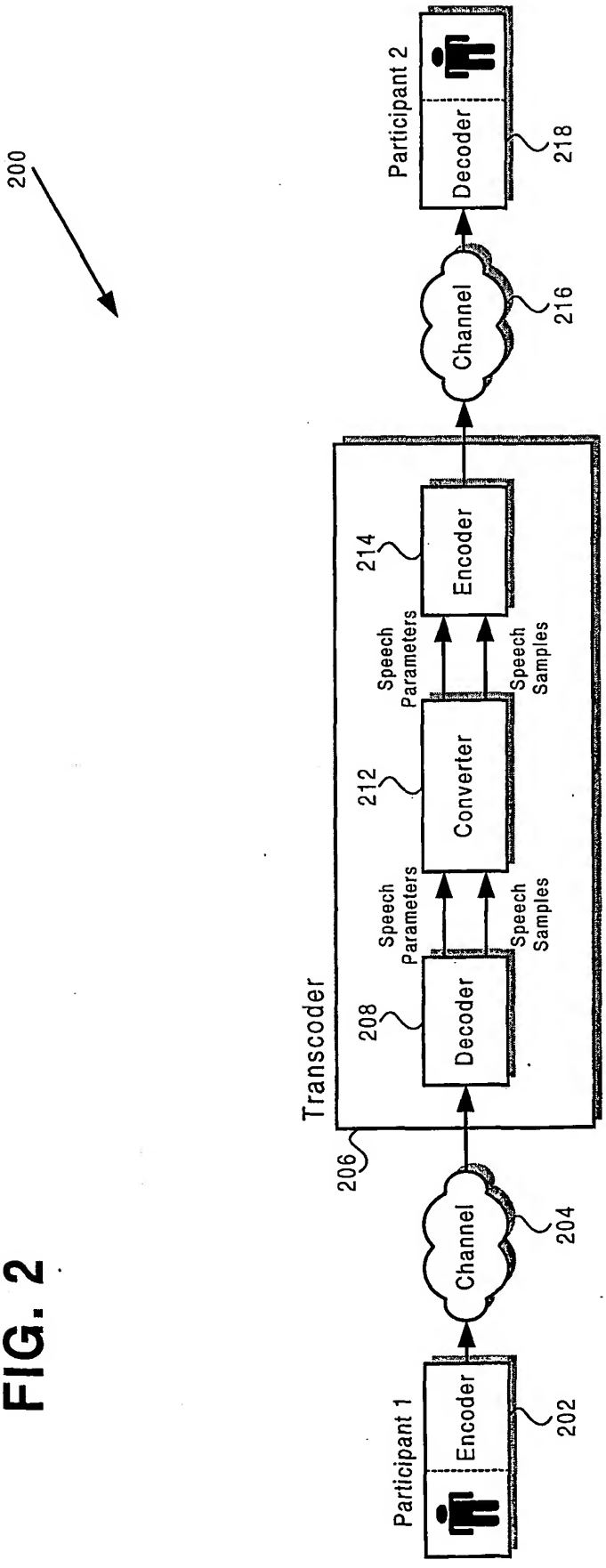

100
Participant 4
106
Gateway
110
Transcoder
Gateway
104
Participant 2
Transcoder
Packet Network
102
Participant 1
Participant 3
108

FIG. 2


200
Participant 1
Encoder
202
Channel
204
Transcoder
206
Decoder
208
Speech Parameters
Speech Samples
Converter
212
Speech Parameters
Speech Samples
Encoder
214
Channel
216
Participant 2
Decoder
218

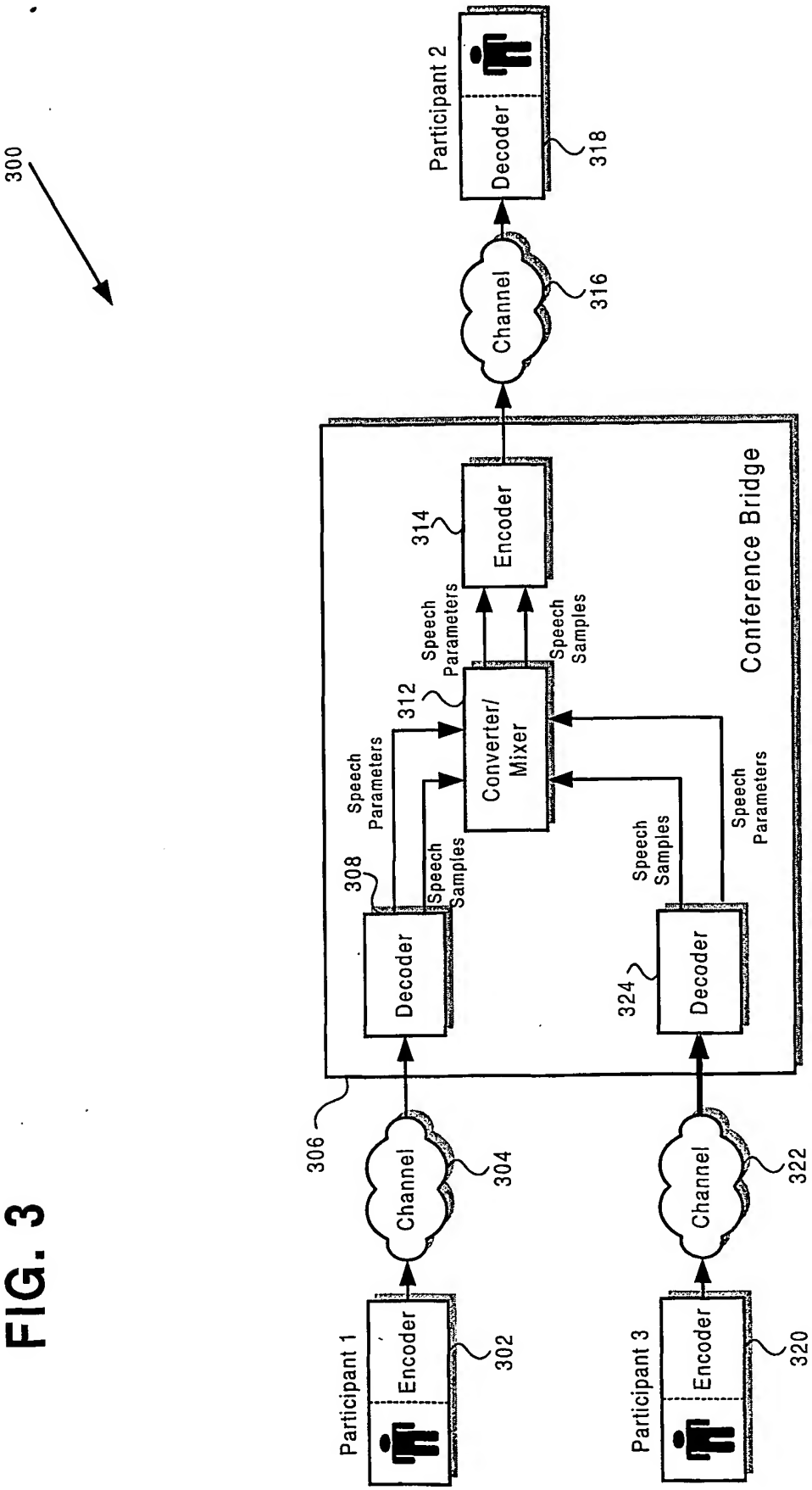

300
Participant 1
Encoder
302
Channel
304
306
Participant 2
Decoder
318
Channel
316
Conference Bridge
Encoder
314
Converter/
Mixer
312
Speech
Parameters
Speech
Samples
Decoder
308
Decoder
324
Participant 3
Encoder
320
Channel
322

FIG. 3

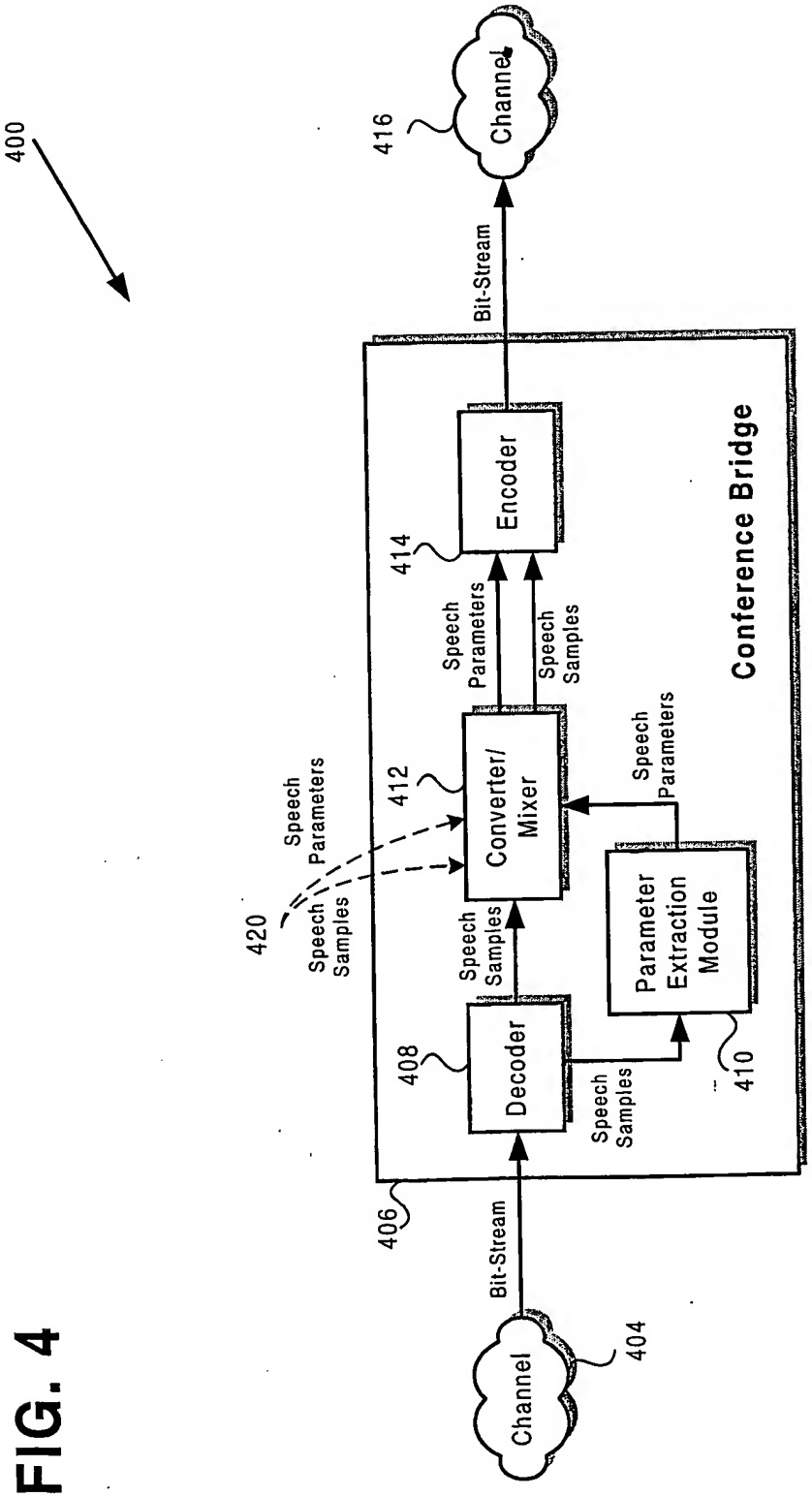
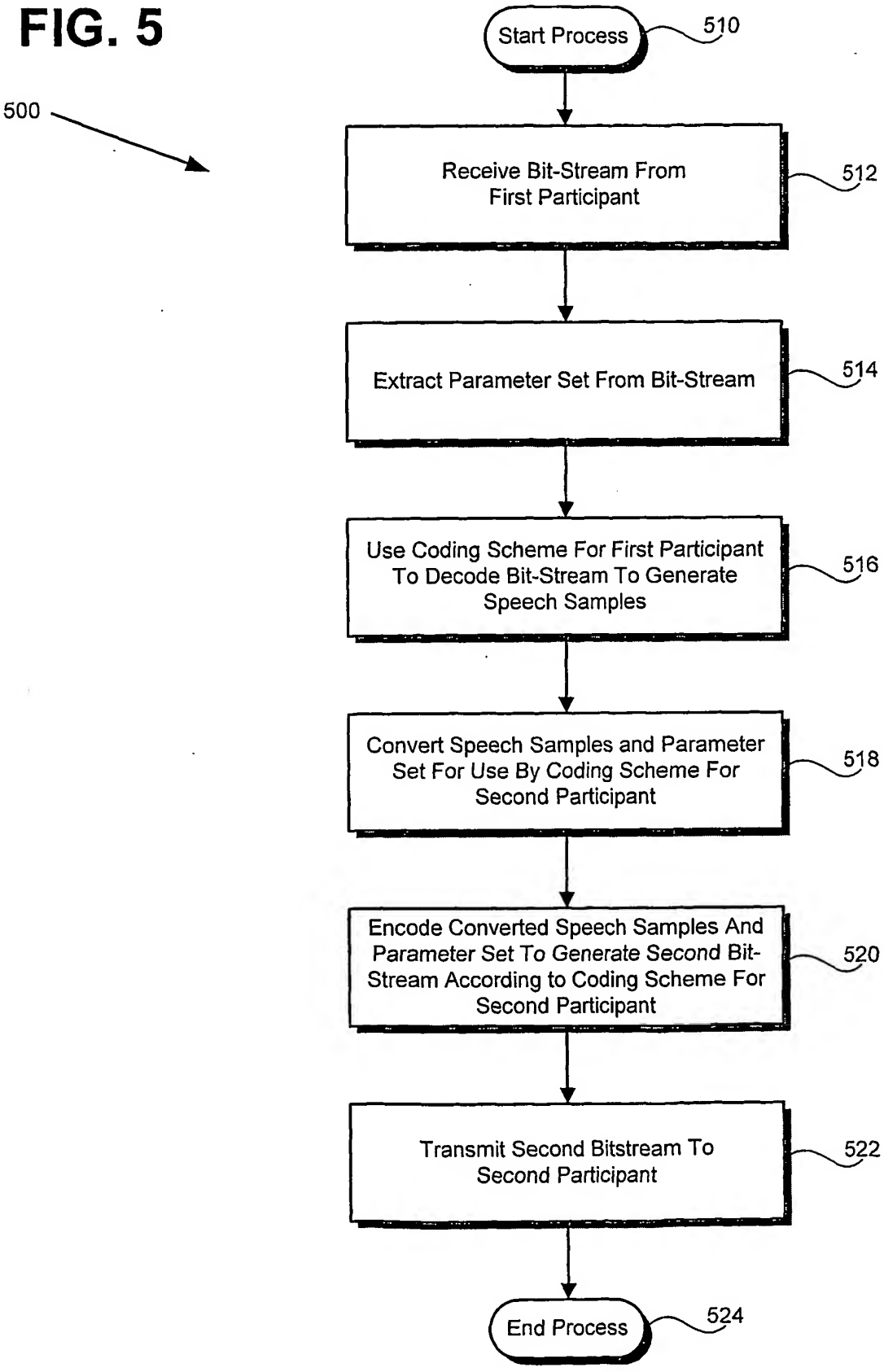

400
Channel
416
Bit-Stream
Encoder
414
Speech Parameters
Speech Samples
Conference Bridge
Converter/ Mixer
412
Speech Parameters
Speech Parameters
Speech Samples
420
Speech Samples
Parameter Extraction Module
410
Decoder
408
Speech Samples
406
Bit-Stream
Channel
404

FIG. 4

FIG. 5


500
Start Process
510
Receive Bit-Stream From First Participant
512
Extract Parameter Set From Bit-Stream
514
Use Coding Scheme For First Participant To Decode Bit-Stream To Generate Speech Samples
516
Convert Speech Samples and Parameter Set For Use By Coding Scheme For Second Participant
518
Encode Converted Speech Samples And Parameter Set To Generate Second Bit-Stream According to Coding Scheme For Second Participant
520
Transmit Second Bitstream To Second Participant
522
End Process
524

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US03/06335

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : G10L 11/04, 11/00, 21/00, 19/00
US CL : 704/ 270.1, 207, 270, 500
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 704/ 270.1, 207, 270, 500, 201

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 5,497,396 A (DELPRAT) 05 MARCH 1996 (05.03.1996), see entire document	1-3, 10-13, 20-22, 29-31 --- 4-9, 14-19, 23-28
X --- Y	US 5,771,452 A (HANLEY ET AL) 23 JUNE 1998 (23.06.1998), Abstract; Figures 4A-4B; column 6, line 29 continuing to col. 7, line 21	1-3, 10-13, 20-22, 29-31 --- 4-9, 14-19, 23-28
Y	US 6,006,178 A (TAUMI ET AL) 21 DECEMBER 1999 (21.12.1999), col. 1, line 56 continuing to col. 2, line 20; column 2, line 50 continuing to col. 4, line 25	6-9,16-19, 25-28

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
19 May 2003 (19.05.2003)

Date of mailing of the international search report
30 OCT 2003

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (703)305-3230

Authorized officer
Marsha Banks-Harold
Telephone No. 703-305-0377

Form PCT/ISA/210 (second sheet) (July 1998)